Presented at IMAC 18, Feb 7-10, 2000, San Antonio, Texas.

UNSUPERVISED LEARNING METHODS FOR VIBRATION-BASED DAMAGE DETECTION

Michael L. Fugate[1], Hoon Sohn[2], and Charles R. Farrar[2]
[1] Computer Research & Applications Group, CIC-3, M/S B265
[2] Engineering Analysis Group, ESA-EA, M/S P946
Los Alamos National Laboratory, Los Alamos, NM 87545

ABSTRACT

The basic premise of vibration-based damage detection is that damage will significantly alter the stiffness, mass, or energy dissipation properties of a system, which, in turn, alter the measured dynamic response of the system. Although the basis for vibration-based damage detection appears intuitive, its actual application poses many significant technical challenges. A fundamental challenge is that in many situations vibration-based damage detection must be performed in an unsupervised learning mode. Here, the term unsupervised learning implies that data from damaged systems are not available. These challenges are supplemented by many practical issues associated with making accurate and repeatable vibration measurements at a limited number of locations on complex structures often operating in adverse environments. This paper will discuss two statistical methods for approaching the unsupervised learning damage detection problem. The first method is density estimation and significance testing. The second method is statistical process control. Examples of these methods are applied to data from an undamaged and subsequently damaged concrete column.

1. INTRODUCTION

The process of implementing a damage detection strategy is often referred to as *structural health monitoring*. This process involves the definition of potential damage scenarios for the system, the observation of the system over a period of time using periodically spaced measurements, the extraction of features from these measurements, and the analysis of these features to determine the current state of health of the system. The output of this process is periodically updated information regarding the ability of the system to continue to perform its desired function in light of the inevitable aging and degradation resulting from the operational environments. Many local damage detection methods have been developed and are routinely applied to a variety of structures [1]. Doebling, et al. [2], present a recent thorough review of more global vibration-based damage identification methods. While the references cited in this review propose many different methods for extracting damage-sensitive features from vibration measurements, few of the cited references take a statistical approach to quantifying the observed changes in these features.

This paper will first pose the general problem of the structural health monitoring in the context of a problem in statistical pattern recognition. For most applications of vibration-based damage detection to large structural systems, data will not be available from a damaged system. Therefore, the pattern recognition must be performed in an unsupervised learning mode, which is primarily aimed at identifying statistically significant outliers in a distribution of the damage sensitive features or changes in the distribution itself. The paper will describe two statistical procedures often employed for unsupervised learning problems. These procedures include density estimation and a statistical analysis procedure referred to as statistical process control (SPC). Although SPC is a well-established condition monitoring procedure for rotating machinery [3], the authors are not aware of applications of this technology to the vibration-based damage detection problem for large structural systems.

2. THE STATISTICAL PATTERN RECOGNITION PARADIGM

In the context of statistical pattern recognition the process of vibration-based damage detection can be broken down into four parts: 1. Operational Evaluation, 2. Data Acquisition and Cleansing, 3. Feature Extraction and Data Compression, and 4. Statistical Model Development

Operational evaluation answers four questions in the implementation of a structural health monitoring system: 1. How is damage defined for the system being monitored? 2. What are the conditions, both operational and environmental, under which the system to be monitored functions? 3. What are the limitations on acquiring data in the operational environment?, and 4. What are the economic and/or life safety motives for performing the monitoring? Operational evaluation begins to define why the monitoring is to be done, what will be monitored, when to monitor, where to monitor, how to perform the monitoring as well as tailoring the monitoring to unique aspects of the system and unique features of the damage that is to be detected.

The *data acquisition* portion of the structural health monitoring process involves selecting the types of sensors to be used, the locations where the sensors should be placed, the number of sensors to be used, and the data acquisition/storage/transmittal hardware. Other considerations that must be addressed include how often the data should be collected, how to normalize the data, and how to quantify the variability in the measurement process. *Data cleansing* is the process of selectively choosing data to accept for, or reject from, the feature selection process. Filtering and data decimation are two of the most common methods for data cleansing.

The area of the structural damage detection process that receives the most attention in the technical literature is *feature extraction*. Feature extraction is the process of the identifying damage-sensitive properties derived from the measured vibration response that allows one to distinguish between the undamaged and damaged structure. Many different damage-sensitive features have been proposed including relatively simple ones such as the root-mean squared (RMS) and kurtosis of the response amplitudes [4], basic linear modal properties such as resonant frequencies and mode shape, and quantities derived from them such a dynamic flexibility matrices, [5]. For systems exhibiting nonlinear response, features such as an Aries Intensity Factor [6] features based on time-frequency analysis [7] and features derived from time series analysis [8] have shown promise for damage detection.

The diagnostic measurement needed to perform structural health monitoring typically produces a large amount of data. *Data compression* into feature vectors of small dimension is often necessary if accurate estimates of the feature statistical distribution are to be obtained. The need for low dimensionality in the feature vectors is referred to as the "curse of dimensionality" and is discussed in detail in general texts on statistical pattern recognition [9].

The portion of the structural health monitoring process that has received the least attention in the technical literature is the development of statistical models to enhance the damage detection process. *Statistical model development* is concerned with the implementation of the algorithms that analyze the distribution of extracted features in an effort to determine the damage state of the structure. The algorithms used in statistical model development usually fall into the three general categories: 1. Group Classification, 2. Regression Analysis, and 3. Outlier Detection. The appropriate algorithm to use will depend on the ability to perform *supervised* or *unsupervised* learning. Here, supervised learning refers to the case were examples of data from damaged and undamaged structures are available. Unsupervised learning refers to the case were data is only available from the undamaged structure. The focus of this paper will be on unsupervised methods.

The statistical models are typically used to answer a series of questions regarding the presence, location, and type of damage. The statistical models are also used to minimize errors, which may be of two types: 1. *False-positive* damage indication (indication of damage when none is present), and 2. *False-negative* damage indications (no indication of damage when damage is present).

3. SUPERVISED AND UNSUPERVISED LEARNING

In this section two approaches of statistical pattern recognition are discussed, supervised and unsupervised learning. The particular application of statistical pattern recognition we have in mind is the detection of damage in a large structural system; for example, a bridge or an aerospace structure. Some of the material in this section and the next section parallels that presented in [10] and [11]. The reader is referred to these two books for a more thorough discussion.

The aim of supervised learning is to classify an object to one of k predefined and mutually exclusive categories based on some rule or rules. Unsupervised learning is often referred to as cluster analysis or unsupervised clustering. The goal of unsupervised classification is to define classes based on a collection of objects.

To illustrate supervised classification, suppose that every object can be described in terms of a feature vector, which is often numerical. When taken together the collection of feature vectors spans a multivariate space, called the feature space. Each object has a true class, and the objective is to construct a rule that assigns the objects to their true class. Typically, the true class for all objects is unknown but the true class for a sample of objects is known. This sample is referred to as a training set or a learning set. The training set is used to construct a classification rule that is then used to predict the class of new objects, based only on their feature vectors. In contrast, for unsupervised learning the true class of the training set objects is unknown and the problem is to define the true classes.

If the training set is a random sample from some population, then regions of the feature space that are densely populated by feature vectors from class k and sparsely populated by feature vectors from other classes, should lead to new observations in that region being assigned to class k. This result suggests that a new feature vector, x, be assigned to class k if the probability of class k given x is greater than the probability of any other class given x. This conditional probability is written $f(k \mid x)$ with estimated value $\hat{f}(k \mid x)$. Note that f may also represent a conditional density function.

For supervised classification to work well, the distribution of feature vectors seen in the future must be the same as the distribution of feature vectors in the training set. If the population that the future objects are drawn from drifts away from the population that the training set was drawn from, supervised classification cannot be expected to work well. In that case, we would be trying to classify new objects to classes that are no longer appropriate.

Applying supervised classification techniques to damage detection in large structural systems requires the construction of a feature vector whose distribution changes when the structure is damaged. Also, training data is needed from the possible damage classes. It should not be difficult to obtain training data from a structure that is in good condition, but getting examples of data from the same structure when it is damaged may be difficult.

One way to proceed is to estimate the density of feature vectors taken when the structure is in good condition and then quantify how consistent future data is with this estimated density. This unsupervised learning approach and how it might be implemented is discussed in more detail in the following sections.

4. DENSITY ESTIMATION APPLIED TO DAMAGE DETECTION

In this section parametric and nonparametric density estimation are discussed. In parametric estimation the observed data are assumed to come from some family of distributions, indexed by the unknown parameters. The observed data are used to estimate the unknown parameters and hence the density. Nonparametric density estimation, by contrast, does not force the data into any particular family of distributions but, instead, allows the data to "choose" the particular density estimate. Several nonparametric density estimation methods are discussed.

4.1 PARAMETRIC DENSITY ESTIMATION

Parametric density estimation is concerned with estimating the unknown parameters of an assumed density function. Two commonly used methods of parameter estimation are least squares and maximum likelihood.

Probably the most familiar example of parametric density is estimating the mean and variance of a normal distribution. Based on a random sample of observations, the mean and variance of the density are estimated by the sample mean and the sample variance of the observed data, respectively.

Estimating a multivariate density is in principle no more difficult than a univariate density; however, there will be more parameters to estimate and therefore generally more data is needed. Estimating a multivariate normal density is common and in this case there is a mean vector and a covariance matrix to estimate, instead of a mean and a variance, as in the univariate example.

4.2 NONPARAMETRIC DENSITY ESTIMATION

In this section some of nonparametric density estimators are reviewed. In particular, a naïve method, a kernel method, and a nearest neighbor method are considered. Throughout this section assume the data, $X_1, \dots, X_n$, are independent and identically distributed observations taken from a continuous distribution with probability density function *f*. The *X*'s are the feature vectors and are initially considered to be univariate.

4.2.1 NAÏVE ESTIMATOR

Nonparametric density estimation can be motivated by observing that, if X is a random variable with density function $f(x)$, then

$$f(x) \approx (1/2h)\Pr(x-h < X < x+h)\,. \tag{1}$$

For a given value of h, the probability of being in the interval $(x-h, x+h)$ can be estimated as the proportion of observed data in the interval. The naïve estimate of the density *f* is

$$\hat{f}(x) = \frac{1}{2hn}\left[\#\text{ of } X_1, \dots, X_n \text{in} \left(x-h, x+h\right)\right] \tag{2}$$

By defining a weight function w as $w(x) = 1/2$ if $|x|<1$ and 0 otherwise, the naïve estimator can be written

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} w\left(\frac{x-X_i}{h}\right). \tag{3}$$

This estimator can be constructed by placing a "box" of width $2h$ and height $(2hn)^{-1}$ over each data point and then adding the heights of all the boxes.

4.2.2 KERNEL ESTIMATOR

A generalization of the naïve estimator is to replace the weight function w by a kernel function K where

$$\int_{-\infty}^{\infty} K(x)\,dx = 1\,. \tag{4}$$

Often, K will be a symmetric density function, for example the normal density function. The kernel estimator with kernel K is defined by

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \tag{5}$$

The parameter h is the window width, also called the smoothing parameter or the bandwidth. Just as the naïve estimator can be thought of as a sum of boxes placed over the observations, the kernel estimator can be thought of as a sum of "hills" placed over the observations.

Two questions naturally come to mind: How do we choose a kernel and how do we choose the bandwidth? Generally, researchers have attempted to answer these questions by considering a global measure of fit, the mean integrated square error (MISE), for an estimator $\hat{f}$ of f. The $MISE$ is defined by

$$MISE(\hat{f}) = E\int\left(\hat{f}(x) - f(x)\right)^2 dx \tag{6}$$

The expectation is with respect to the sampling distribution of $\hat{f}$. Whether or not the $MISE$ is an appropriate measure of global fit depends on the particular problem at hand. A general recommendation concerning the choice of kernels is for cautious users to choose a symmetric non-negative unimodal kernel. In addition, the only kernels that guarantee $\hat{f}$ is everywhere non-negative are kernels that are probability density functions.

The choice of the smoothing parameter is important. If h is too small the data will be "under-smoothed" and random features of the data tend to be emphasized. If h is too large the data will be "over-smoothed" and systematic features will be lost. In particular, if the underlying density is bimodal and h is chosen too large, the bimodal feature will not be seen in $\hat{f}$.

Results have been derived for choosing an optimal bandwidth. Unfortunately, under general conditions the optimal bandwidth depends on the unknown density f. As might be expected, the more rapidly the density fluctuates, the smaller the value of h that is needed. Good results can often be obtained by using a Gaussian kernel and $h = 0.9\,An^{-1/5}$ where

$$A = \min\text{ (standard deviation, interquartile range/1.34)} \qquad (7)$$

The standard deviation and the interquartile range are computed from the observations. At the least, this choice of h should provide a good starting point. A better choice of h might be found by making use of cross-validation techniques.

One modification of the kernel method, referred to as an adaptive kernel estimator and particularly useful in multivariate estimation, is to allow a different bandwidth for each observation. By allowing a different bandwidth the tails of the density can possibly be better fit. Essentially, adaptive kernels attempt to smooth the tail observations more than observations in the center of the density. Smoothing tail observations prevents the estimated density from showing spurious "bumps" in the tail.

4.2.3 NEAREST NEIGHBOR ESTIMATOR

The nearest neighbor method is based on the fact than in a sample of size n, the expected number of observations in the interval $[x-h, x+h]$ is about $n(2h)\,f(x)$, for each $h > 0$. Now, for every x define

$$d_1(x) \le d_2(x) \le \cdots \le d_n(x) \qquad (8)$$

to be the distances, arranged in ascending order, from x to the points of the sample. Exactly k observations will fall in the interval $[x - d_k(x),\ x + d_k(x)]$, so $k \approx 2n\,d_k\ f(x)$. this suggests defining the kth nearest neighbor density estimate to be

$$\hat{f}(x) = \frac{k}{2n\,d_k(x)} \qquad (9)$$

The kth nearest neighbor method allows the size of the box over the observations to change so that each box contains the same number of observations, in this case k. In the tails of the distribution the distance $d_k(x)$ will be larger than in the main part of the distribution, and so the problem of under-smoothing the tails should be reduced.

The nearest neighbor estimate is *not* a density function and the tails of the estimated function tend to die away very slowly. If interest lies in exploring tail behavior, as will be the case in many damage detection applications, then nearest neighbor methods should probably not be used.

Note that one choice of bandwidth for a kernel estimate is to use $h = d_k(x)$. This choice of h allows for a smoother estimate in the tails.

4.2.4 MULTIVARIATE DENSITY ESTIMATION

All of the methods discussed so far can be used with multivariate data. For example, instead of a univariate kernel a multivariate kernel is used. Also, for practical reasons, multivariate kernels are usually restricted to ellipsoidal probability density functions. A common choice is a multivariate Gaussian kernel. Now, instead of a single smoothing parameter h, we will have a smoothing parameter matrix H.

Another choice for a multivariate kernel is the product of d univariate kernels where d is the dimension of the feature vectors. Each univariate kernel can be different and each one can depend on a different smoothing parameter h_j. Typically each kernel is restricted to have the same shape.

In principle, there is no difficulty extending the ideas of univariate density estimation to higher dimensions. However, the nature of high dimensional spaces presents some peculiar problems. In one dimension, nearly 90% of the "mass" of a standard normal density is within 1.6 standard deviations of 0. With a ten-dimensional standard multivariate normal density, nearly 99% of the "mass" will be at a distance greater than 1.6 standard deviations from the origin and, unfortunately, this area of space is sparsely populated.

As might be expected, trying to get a good estimate in the tails of a multivariate density is going to be difficult. Silverman [8] makes the strong statement that it is futile to expect very good estimates in the tails if the dimension of the data is greater than 3 or 4 unless enormous amount of independent data are available.

As an illustration of the problems encountered in high dimension, consider a simple histogram in 10 dimensions, i.e. there are 10 variables. If the range of each variable is divided up into 10 bins then the 10 dimensional histogram will have 10^{10} cells. Few samples will be large enough to get accurate estimates within each cell and in fact most cells will have no observations.

5. DENSITY ESTIMATION APPLIED TO DAMAGE DETECTION

Assume a sensor that measures acceleration has been placed on a structure and that the structure is in good condition. Denote the damage-sensitive feature derived from the acceleration measurements taken at time t by $x(t)$. These features are used to construct a univariate density estimate.

Once the density function has been estimated, the next step is to determine if future data features come from this density. Because the density estimate was constructed using data obtained when the structure was in good condition, if the future data isn't consistent with the density estimate, this is *evidence* that the structure may be damaged. The important issue now is how to determine if future data is consistent with the estimated density.

A natural way to proceed is to consider the new feature to be consistent with the density if the feature is not too far from the center of the density. For example, if a new feature is below the 1st percentile or above the 99th percentile of the density, this result would be very unusual if in fact the new feature is generated by the same process that generated the data when the structure was in good condition.

If instead of a single sensor there were multiple sensors on the structure, the observations, $x(1), \ldots, x(n)$, would be multivariate and the density estimate would be multivariate.

For a multivariate density contours of constant density can be constructed. A new data feature might be considered unusual if it is beyond, say, the 99th percent contour, i.e. the contour such that 99 percent of the mass of the density is inside the contour.

Even if the new feature is beyond the 99th percent contour, it may be the result of changing operational or environmental conditions as opposed to damage. Therefore, it is necessary that the feature vector contain parameters that quantify variability in the operational and environmental conditions as well as parameters that quantify change in the structural condition. The tradeoff will be that quantifying the operation and environmental conditions necessitates increasing the dimension of the feature vector with the associated dimensionality problems previously discussed.

6. STATISTICAL PROCESS CONTROL

In this section statistical process control (SPC) [12] and how it can be applied to vibration-based damage detection is discussed. Again, suppose at some location on a structure there is one sensor for recording acceleration and denote the features derived from the measurement at time t by $x(t)$. After determining that the structure is in good condition there will be an almost continuous stream of acceleration measurements. Based on these observed features, one would like to know if the structure is in good condition, or if it has been damaged.

Features derived from measurements taken when the structure is in good condition will have some distribution with mean m and variance s^2. If the structure is damaged, there might be a change in the mean, the variance, or both.

Statistical process control provides a framework for monitoring future features and for identifying new data that is inconsistent with past data. In particular, quality control charts are proposed to monitor the mean, the variance, or some other function of the features derived from the acceleration measurements.

If the mean m and standard deviation s are known, a control chart is constructed by drawing a horizontal line at m and two more horizontal lines representing the upper and lower control limits. The upper limit is at $m + ks$ and the lower limit is at $m - ks$. The number k is chosen so that when the structure is in good condition a large percentage of the observations will fall between the control limits. Often k is chosen so that at least 99% of the observations are between the limits.

As each new measurement is taken, the new feature is plotted versus time. If the condition of the structure has not changed, almost all of these features should fall between the upper and lower control limits, the exact percentage being determined by the choice of k. In addition, there should be no obvious pattern in the charted data; e.g., there should not be a repeated pattern of 5 observations above the mean followed by 5 observations below the mean. If the structure is damaged there might be a shift in the feature mean, which would be indicated by an unusual number of charted values beyond the control limits. Plotting the individual measurements on a control chart is referred to as an X–chart, or a Shewhart chart.

To detect changes in the mean of the features, an intuitively appealing idea is to form rational subgroups of size *n*, compute the sample mean within each subgroup and chart the sample means. The centerline of the control chart would still be m but the standard deviation of the charted values would be $s/\sqrt{n}$ so the control limits would be placed at $m \pm ks/\sqrt{n}$. This type of control chart is referred to as an X-bar chart.

The subgroup size n is chosen so that observations within each group are, in some sense, more similar than observations between groups. If n is chosen too large a drift that may be present in the mean can possibly be obscured, or averaged-out. An additional motivation for charting sample means, as opposed to individual observations, is that the distribution of the sample means can be approximated by a normal distribution.

When the mean and variance are unknown they must be estimated from observed data, taken when the structure is in good condition. The mean could be estimated with the sample mean of the feature. Several different methods have been proposed for estimating s. For example, the sample standard deviation or some function of the range of the data could be used. Alternatively, if rational subgroups are constructed, the sample standard deviation or a function of the range within each subgroup could be computed and then the estimates pooled across the groups. Montgomery [12] has a more complete discussion of estimating a standard deviation for use in control charts.

This discussion of control charts has assumed that the features are uncorrelated. In practice the observed data likely to be autocorrelated. When there is autocorrelation in the data, the control limits as just described are inappropriate because the estimate of s is inappropriate. One way of removing autocorrelation is to fit an autoregressive model to the data. If the model is approximately correct, the residuals from the fitted model should be nearly uncorrelated. A control chart can then be constructed, exactly as previously described, using the residuals as the data feature.

7. EXAMPLES

In this section we illustrate kernel density estimation and an X-bar chart. For a more complete discussion of the data used in section and the experiment that generated the data, see [13].

Acceleration measurements were taken on an undamaged in-situ bridge column at 40 sensor locations. The column was then damaged and acceleration measurements taken at each sensor location. In all there were 5 damage levels and one undamaged level that is designated damage level 0. Only the measurements from sensor location 1 are considered in this section.

At each damage level there is a time-history of 8192 acceleration measurements. An autoregressive model was fitted to the damage level 0 data and the residuals computed; for the density estimates an order 3 model was fit and for the control chart an order 5 model was fit. The autoregressive model estimated from the damage level 0 data was then fitted to the data from damage levels 1 through 5 and the residuals computed. The residuals from damage levels 0 through 5 are the data for the examples in this section.

Figure 1 shows a kernel density estimate for each damage level. The normal density was chosen as the kernel function with smoothing parameter of approximately 6×10^{-5}. The estimated density functions for damage levels 1 through 5 are all clearly different from damage level 0. The most notable feature of the estimated densities is the increase in variability.

Figure 2 shows estimated densities obtained by splitting the damage level 0 in half. The kernel function was a normal density with smoothing parameter 1.0×10^{-4}.

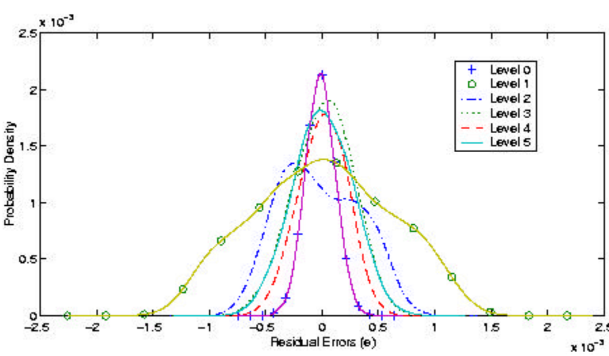

Figure 1: Kernel density estimation of residual errors

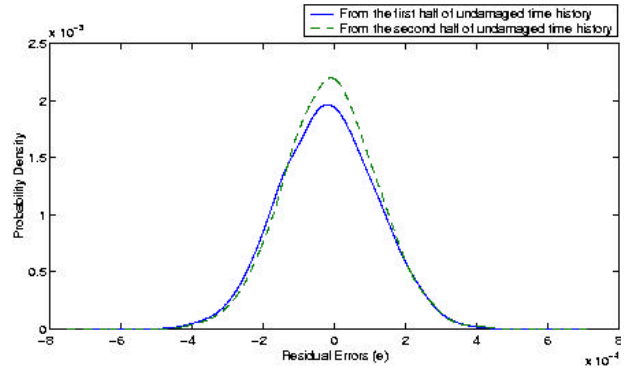

Figure 2: Comparison of density functions of the residuals computed from two segments of the undamaged time history

To construct the control charts in Figure 3, rational subgroups of size 4 were created by placing the first 4 residuals in group 1, the next 4 in group 2 and so forth; this gives 2046 subgroups. The sample mean was computed for each group and then charted. The upper and lower control limits were set so that approximately 99% of the charted values should fall between the limits, when there is no damage.

Figure 3a shows the charted values computed from damage level 0 data and Figure 3b shows the charted values computed from damage level 1 data. In Figure 3a there are 13 values outside the limits. If there was no damage, then Figure 3b should show about as many values beyond the control limits as Figure 3a. However, in Figure 3b there are 399 charted values beyond the control limits.

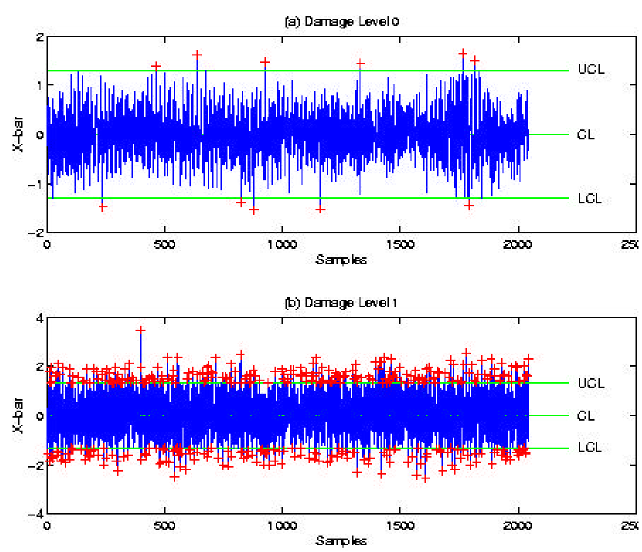

Fig. 3 Control charts for undamaged column (top) and column after the first level of damage (bottom).

8. SUMMARY

This paper has described the general process of structural health monitoring as in terms of a problem in statistical pattern recognition. Because most structural health monitoring applications to large structural systems do not allow the opportunity to obtain data from the damaged system, statistical pattern recognition will have to be applied in an unsupervised learning mode. This paper has attempted to summarize various statistical procedures that can be employed for such unsupervised learning problems.

The two general unsupervised-learning statistical procedures described in this paper, (1) density estimation & significance testing, and (2) statistical process control, are aimed at identifying statistically significant outliers in a distribution of the damage sensitive features or change in the distribution itself.

Two key issues arise when attempting to apply the statistical procedures described in this paper to structural health monitoring. First, one must examine features that quantify environmental and operational variability as well as the structural condition. Without such measures, one will have difficulty distinguishing changes in vibration response caused by damage from changes caused by other sources of variability. Second, the addition of these environmental and operational features must be done with the concept of feature vector dimensionality in mind. Particular attention must be paid to the difficulties that large dimensional feature vectors pose for statistical quantification.

Finally, no matter which features are used and which statistical pattern recognition method is employed, there is always the need to perform false-positive studies as shown in the example presented herein.

ACKNOWLEDGEMENTS

We are grateful to George Papcun of Computer Research & Applications Group for his insight that he shared with us and for his many helpful suggestions. Portion of the funding for this work has come from a cooperative research and development agreement with Kinemetrics Corporation, Pasadena California. The civilian application of this CRADA is aimed at developing structural health monitoring systems for civil engineering infrastructure. A portion of the funding for this research was provided by the Department of Energy's Enhanced Surveillance Program. Vibration tests on the columns were conducted jointly with Prof. Gerry Pardoen at the University of California, Irvine using Los Alamos National Laboratory's University of California Interaction Funds. CALTRANS provided the primary funds for construction and cyclic testing of the columns.

REFERENCES

[1] Doherty, J. E., "Nondestructive Evaluation," Chapter 12 in *Handbook on Experimental Mechanics*, A. S. Kobayashi Edt., Society for Experimental Mechanics, Inc., 1987

[2] Doebling, S. W., C. R. Farrar, M. B. Prime and D. W. Shevitz, (1998) "A Review of Damage Identification Methods that Examine Changes in Dynamic Properties," *Shock and Vibration Digest* **30** (2).

[3] Wowk, V., (1991) *Machinery Vibration Measurement and Analysis*, McGraw-Hill, Inc. New York, 1991.

[4] Martin, H. R., (1989) "Statistical Moment Analysis as a Means of Surface Damage Detection", *7th IMAC*, pp. 1016-1021.

[5] Farrar, C. R. and D. V. Jauregui (1998) "A Comparative Study of Damage Identification Algorithms Applied to a Bridge: Part I, Experimental" *J. of Smart Materials and Structure,* **7**.

[6] Straser, E. G. (1998) "A Modular, Wireless Damage Monitoring System For Structures," Ph. D. Dissertation, Dept. of Civil Eng., Stanford Univ., Palo Alto, CA.

[7] Prime, M. B. and D. W. Shevitz (1996) "Linear and Nonlinear Methods for Detecting Cracks in Beams," 14th *IMAC*, Dearborn, MI.

[8] Hunter, N. F. (1999) "Bilinear System Characterization from Nonlinear Time Series Analysis," 17^{th} *IMAC*, Orlando, FL.

[9] Bishop, C. M. (1995) *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, Uk.

[10] Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall.

[11] Hand, D. J., (1997) *Construction and Assessment of Classification Rules*, John Wiley & Sons Inc.

[12] Montgomery, D. C. (1996) *Introduction to Statistical Quality Control* 3^{rd} ed., John Wiley & Sons, Inc., New York.

[13] Sohn, H., M. L. Fugate and C. R. Farrar, (2000) *"Continuous Structural Monitoring Using Statistical Process Control"*, 18^{th} *IMAC*, San Antonio, TX.